# Choice of Aircraft Fleets in the U.S. Domestic Scheduled Air Transportation System: Findings from a Multinomial Logit Analysis

*The National Airspace System (NAS) in the United States had an inventory of 5,156 big jets at the end of December 2002, of which 4,085 were narrow bodies, and 1,071 were wide bodies. In addition, there were 1,180 regional jets and 660 turboprops in the system at that time. Empirical research reveals that there is a critical link between the flow of scheduled passenger services and the choice of aircraft used by the airlines in serving market pair demand. This relationship can be empirically retrieved without detailed knowledge of airlines' behavior and used for analyzing traffic patterns in the NAS. Using the T100 segment data from the first two quarters of 2004, a multinomial qualitative choice model is developed in this paper. This framework establishes empirical linkages among aircraft choice, and passenger flows, distance, types of airport hubs, network and time of the year. Estimated models demonstrate that both passengers and distance play important roles in selecting types of aircraft. Overall, the model is capable of predicting exact choices 51% of the time; with some flexibility of making a one-off mistake, the model is capable of making almost nine out of 10 choices correctly. Using the estimated coefficients from the qualitative choice model and varying assumptions (number of passengers, in particular), forecasts of aircraft operations by market segments and the fleet mix can easily be generated. These forecasts can then be used to understand the performance of the U.S. NAS.*

## by Dipasis Bhadra

## INTRODUCTION

The U.S. airline network is vast. About 36,000 origin and destination markets are currently served by more than 50,000 flight segments (DOT 2004). Scheduled air carriers transport more than a million passengers undertaking around 15,000 departures a day (Air Transport Association 2004). This scale of operations is unprecedented in the history of aviation. Scheduling aircraft in a network of this magnitude is understandably a complex task.

To make matters even more challenging, the majority of the U.S. airlines maintain a heterogeneous fleet structure (Figure 1). Determining aircraft choice for flight segments is indeed a challenging operational task. Airlines make these choices on a daily basis. In fact, rational choices of aircraft fleets in serving both origin and destination markets and associated segments[1] are an essential part of airlines profit optimization.[2] Airlines spend considerable resources to match passenger

demand with the right size of aircraft, taking into consideration distance, types of airports, connection possibilities, and network.

In this paper, a simple empirical framework demonstrating this choice is introduced. In particular, an attempt is made to answer the following questions: How does the passenger demand by segment of flight influence the choice of aircraft? How does the length of the haul affect the choice of aircraft? Can the choice of aircraft and fleet mix be derived from the knowledge of demand for scheduled air services? By examining these relationships empirically, one can uncover the complexities of aircraft scheduling by flight segments. In other words, an attempt has been made to fill the existing information void by offering an explicit choice mechanism that may be used to convert passengers into aircraft choice by segment pairs for domestic U.S. scheduled air services. This information may add to the knowledge on how aircraft can *actually*[3] be scheduled by the airlines. The forecasts of passengers and other

**Figure 1: 2002 Distribution of Aircraft: American Airlines**
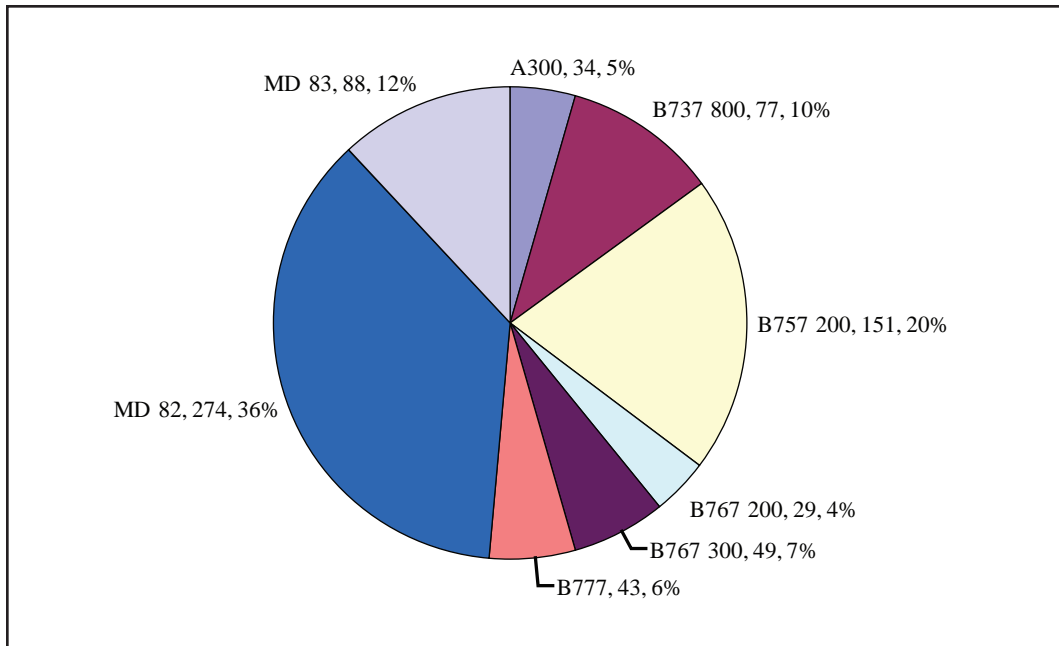


Source: Airline Monitor (2003)

exogenous variables along with their estimated parameters can also be used to generate forecasts of aircraft operations. Understanding airline operations is mandatory for analyzing national air space system (NAS) performance. A large group of simulation models (Bhadra, et. al. 2005) that use different assumptions regarding aircraft operations presently exist. All of them convert passengers into aircraft operations by making specific assumptions regarding the load factors and the size of the aircraft. None of them make use, to the best of our knowledge, of any choice mechanism to include passenger flows and other factors, including distance, in determining the aircraft fleet choice.

The paper is organized as follows: Section II provides a background for the paper including a brief survey of literature. Section III introduces the conceptual and empirical framework for analyzing the choice of aircraft. Section IV describes the data and their sources. Section V reports results and provides a comparison between the observed choices (i.e., choices that have actually been made) and choices indicated by the empirical model (i.e., probabilistic choices). Section VI concludes the paper.

**BACKGROUND AND LITERATURE REVIEW**

Broadly categorized, aggregate fleet structure in the country has a definite pattern: about 58% of the aircraft in the national air space system (NAS) is a narrow body aircraft while about 32% is either a wide body or a regional jet (Figure 2).

However, within those broad categories, the fleet structure is diversified. Diversifying the aircraft fleet has its own trade-offs. Locking into one type of aircraft may reduce costs (maintenance, inventory, pilot training, and scheduling) considerably, while diversification may allow more choices over both routes served and scheduling. Many of these choices can be captured by specifying operating profit as the objective function for airlines instead of previously used revenue maximization or cost minimization (Fan 2002). The fleet decision in the earlier studies prior to Fan (2002) was fixed, and hence, any incremental gains from revenue were not discounted appropriately by increases in costs. The framework proposed by Fan (2002) recast the airlines optimization problem

in the context of operating profit and, when combined with unconstrained aircraft capacity and fleet size, allowed far more flexibility to the airlines than those under alternative specifications. Consequently, airlines can select both the aircraft size and the groups of targeted passengers and, in general, tend to choose relatively smaller aircraft size for optimizing profit (Fan 2002).

The diversification of aircraft types has associated costs arising from scheduling complexities. Many low-cost airlines have reduced or eliminated this heterogeneity by choosing only one type of aircraft. For example, Southwest Airlines maintains an all B-737 fleet structure while JetBlue has maintained, until recently, an all A320 fleet structure. Even within the broad homogeneous class, the 737 in the case of Southwest Airlines, choosing a particular aircraft can be a complicated task.[4] The degree of these complexities is compounded by the increased number of aircraft choices, markets and segments served. For example, each of the six major airlines[5] maintains a heterogeneous fleet structure. All these airlines also have hubbing operations with varying degrees. The choice of aircraft, under these circumstances, will certainly be far

more complicated compared to an airline with primarily point-to-point service and somewhat homogenous fleet structure (e.g., Southwest).[6]

It is important to notice that the choice of aircraft for domestic travel has been remarkably stable in the past. For example, 737 narrow bodies (i.e., 737-300) have been the top choice throughout the later part of the last decade with a share of almost 18% to 20% of total enplanement. In contrast, the B-727, B-737-1/2, and MD-80 have been going out of the system at a fast rate.

Generally speaking, aircraft are chosen in a particular segment, *ceteris paribus*, to serve passenger needs given other market conditions such as fare and competition. It is obvious that a wide body may not be chosen in a segment where a narrower plane or even a regional jet can adequately meet the existing demand at prevailing fares. Moreover, the techno-economic requirements of an aircraft, e.g., cruise speed, average haul, and associated economics, will also have a bearing on the type of aircraft that will be chosen in a particular segment. For example, wide bodies tend to have relatively lower cost per available seat mile (a standardized measure for cost) than their counterpart, narrow bodies. Broadly speaking,

**Figure 2: Types of Aircraft in the U.S., 2003**
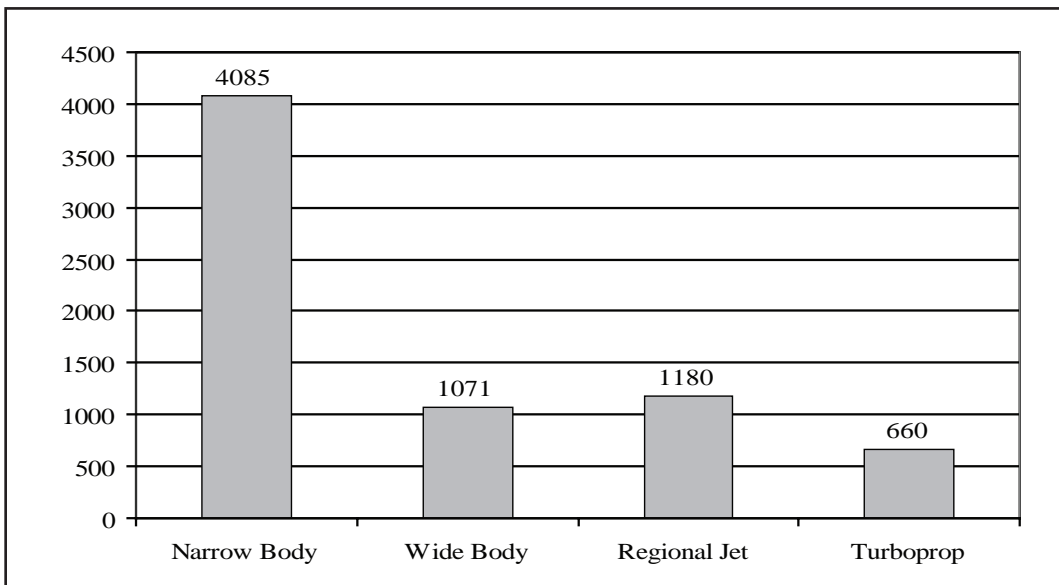


Source: SpeedNews (2003)

aircraft choice is determined by three factors: average flight stage length, average speed, and average number of seats.

Wide bodies, on average, fly longer hauls (2350 miles) compared to narrow bodies (660 miles) at a speed comparatively higher (450 miles per hour compared to 330 miles per hour) and with a higher number of seats (230 seats compared to 130 seats).

These factors together result in cost differentials (Figure 3) and determine the type of aircraft that will be chosen for a particular segment. In addition, given a vast inventory of aircraft for most of the large airlines and those presently parked at the Mojave Desert in California, it is obvious that scheduling of maintenance also plays a crucial role while scheduling for routes.

Deciding on which aircraft to order and the flexibility to change the sizes has substantial financial implications for airlines. Consequently, they take considerable caution in the decision process and often maintain some flexibility in changing the size if demand and market conditions change. Nevertheless, a flexible investment strategy that takes into account long lead times, large expenditure (as in cases of aircraft purchases) and market and

other uncertainties could prove very useful in evaluating these costly investments. Miller and Clarke (2005) demonstrated that such risks can be managed through well-specified real option strategies where investment can benefit from upside turn of events while protecting against downside losses.

Yu (2003) examined the decision process of flight operations using a nested logit model. The choice of aircraft is exogenous to the performance of flight operations (i.e., cancellations and delays) in this study. The study uses four sets of explanatory variables: logistical variables that take into account performance characteristics; competition variables using market and hub characteristics; weather variables that consider effects of severe weather; and aircraft characteristics using number of seats, age and make of the aircraft. The study finds that routes with higher daily frequencies, e.g., hub-and-spoke routes, are subject to more cancellations. As expected, congestion at busy airports adversely affects the on-time performance of flights. Contrary to earlier findings, however, Yu (2003) finds that lack of competition does not necessarily lead to worse service quality (in terms of cancellations and delays).

**Figure 3: Block Hour Operating Cost by Aircraft Type**
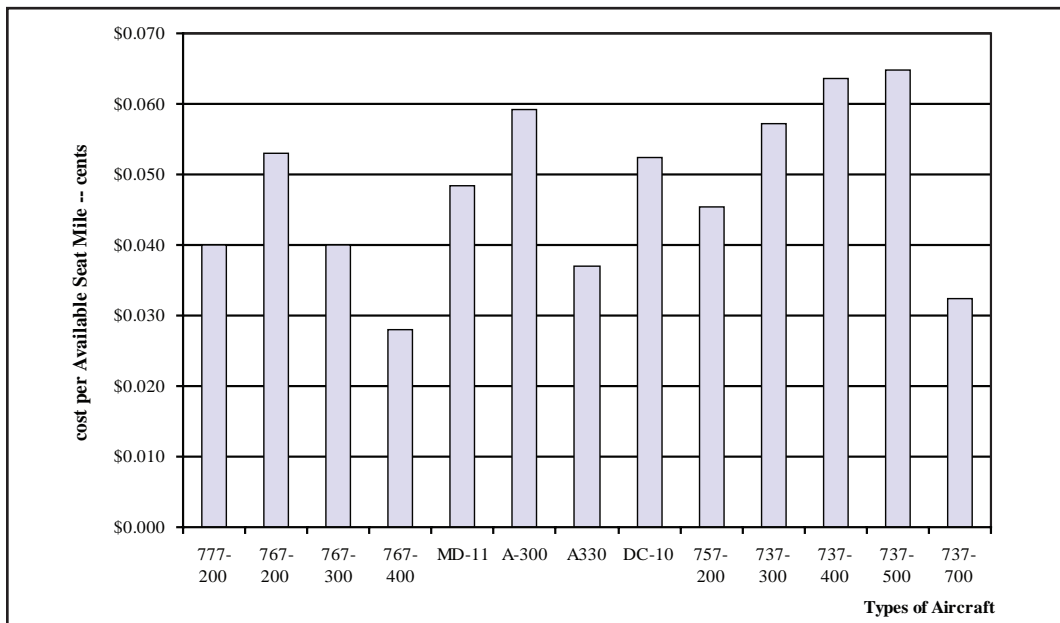


Source: Airline Monitor (2003)

## DETERMINING CHOICE OF AIRCRAFT: THE FRAMEWORK

### Theoretical Framework

For simplicity, let us assume that the representative airline chooses only two-classes of aircraft (i, j). Then, the individual airline's profit maximization can be stated as:

(1)  $\pi$   $p_i$ pax  $c_k$  R          $\forall$ i , j

where $p_i$ is the price per passenger in aircraft i, pax is the number of passengers on board, $c_k$ is the cost of using the aircraft and R captures a host of other factors that are considered to be fixed, e.g., stage length, types of network, types of origin and destination airports, and season.

The first order condition for profit maximization (with respect to aircraft choice) yields:

(2)   $p \,(\text{pax})' - c'_k = 0$

where $p \,(\text{pax})'$ is the marginal revenue and $c'_k$ is the marginal cost.

In general, aircraft should be employed to the point where marginal revenue equals marginal cost (constant for both choices). This condition then can be used to derive the optimal demand condition for aircraft i and j. For aircraft i to be chosen, the indirect profit[7] from i > indirect profit from j or,

(3)   $\Pi_i \,(p_i c'_k; R) > \Pi_j \,(p_j c'_k; R)$

The choice will find an equilibrium for the airline at the margin when $\Pi_I = \Pi_j$.

The problem of choosing one aircraft over another (equation 3) can be transformed into a marginal framework by specifying the discrete choice probabilities as the dependent variable (McFadden 1974). Assume that the profits in the ith and jth aircraft can be best approximated by the following linear function in exogenous variables, X:

(4)   $I = \sum_{k=1}^{K} a_{ki} X_k + v_i$

and,

(5)   $J = \sum_{k=1}^{K} a_{kj} X_k + v_j$

where j = 1,… i…n choices of aircraft; $a_{ki}$ and $a_{kj}$ are the linear coefficients; k = 1….K exogenous variables, and v represents error terms.

Let's define,

(6)   $Y_i^* = I - J$

or,   $= \sum_{k=1}^{K} (a_{ki} - a_{kj}) X_k + (v_i - v_j)$

which can be written as,

(7)   $Y_i^* = \sum_{k=1}^{K} b_{ik} X_k - \mu_i$

where $b_{ik} = (a_{ki} - a_{kj})$; $\mu_i = (v_i - v_j)$. If $Y_i^* = I - J$ > 0 (implying $\mu_i < \Sigma\, b_{ik} X_k$), aircraft i is chosen and if $Y_i^* = I - J < 0$ (implying $\mu_i > \Sigma\, b_{ik} X_k$), then aircraft j is chosen. This allows us to write the probabilistic statement as the following:

(8)   $P_i \,(Y_i = i) \; = P_i \,(Y_i^* > 0) = P \,(\mu_i < \sum_{k=1}^{K} b_{ik} X_k)$

If $u_i$ is a continuous variable and has logistic distribution, and $Y_i$ is binary, then the above choice can be written as a binary logit model; if $Y_i$ has multiple categories then the above equation can be written as a multinomial logit model.

### Binary Logit

In situations when airlines have only two choices of aircraft to assign to flight segments, there is essentially a binary qualitative choice.[8] Because the linear probability model does not guarantee the predicted values of that choice to lie between (0, 1), it requires a process of translating the values of the attribute X (i.e., vector containing explanatory variables explaining the choice) to a probability which ranges in value from 0 to 1. We would also like to maintain the property that such a transformation would allow increases in X to be associated with an increase in the dependent choice variable for all values of X. Together, these requirements suggest the use of the cumulative probability function (F). F is defined as having its value equal to the probability that an observed value of a variable (for every X) will be less than or equal to a particular X. The range of F is then (0, 1) since all probabilities lie between 0 and 1. The resulting probability distribution may be expressed as follows:

(9) $P_i = F \,(\alpha + \beta X_i) = F \,(Z_i)$

where $\alpha$ and $\beta$ are the parameters of the model

and F representing the distribution. Common models in this category include Probit (standard normal), Logit (logistic), and Gompit (extreme value) specifications for the F function. The two cumulative probability functions, the normal (Probit) and the logistic, have been used widely in the literature and among practitioners (see McFadden 1974, Hosner and Lemeshow 1989).

To understand the logit specification, assume that there exists a theoretical continuous index $Z_i$ which is determined by an explanatory variable X. Thus, we can write,

(10) $Z_i = \alpha + \beta X_i$

Observations on $Z_i$ are not available unless data are available that distinguish whether individual observations are in one category (e.g, Aircraft Category 1) or a second category (e.g., Aircraft Category 2). Logit methodology allows solution of the problem of how to obtain estimates for the parameters while at the same time obtaining information about the underlying index Z.

Let Y represent a dummy variable that equals 1 when the Aircraft Category 1 is chosen and 0 when the other category is chosen.[9] Then assume that each individual choice $Z_i^*$ represents the critical cutoff value which translates the underlying index into a choice decision, such as,

Category 1 = 1 if $Z_i > Z_i^*$

(11) Individual choice for

Non-category 1 = 0 if $Z_i \leq Z_i^*$

In this case, the threshold is set to zero, but the choice of a threshold value is irrelevant as long as a constant term is included in $X_i$. The logit model assumes that $Z_i^*$ is a cumulative distribution function for the logistic distribution, so that the probability that $Z_i^*$ is less than (or equal to) $Z_i$ can be computed from the probability distribution function. The standardized cumulative probability distribution function for the logistic distribution is written as:

(12) $P_i\left(y_i = 1 | x_i, \beta\right) = \dfrac{e^{\mu\beta' x_{in}}}{e^{\mu\beta' x_{in}} + e^{\mu\beta' x_{jn}}} = \dfrac{1}{1 + e^{-\mu\beta'\left(x_{in} - x_{jn}\right)}}$

where $x_{in}$ and $x_{jn}$ are vectors describing the attributes of alternatives i (Category = 1) and

j (Category = 0); μ is a scale parameter that is positive in value. When the parameters of $Z_i$ are linear, the parameter μ cannot be distinguished from the overall scale of the β's (Ben Akiva and Lerman, 1984). Often, μ is assumed to be equal to 1. By construction, the variable $P_i$ will lie between (0,1). $P_i$ is the probability that an event occurs, i.e., probability of the choice of Category 1 aircraft.[10]

## Standard Multinomial Logit Model (MNLM)

Often, choices are not restricted to a binary set. As in the case with airlines, choices of aircraft for assignment in flight segment/s are often numerous. Except for a few (e.g., Southwest Airlines), the majority of U.S. airlines have numerous choices of aircraft. Under such circumstances, the choice set will have to be expanded into multinomial choices. Thus, when there is more than one aircraft choice, i.e., category of aircraft (A/C) = 1,…,J, then, the probability is associated with all those choices are $P_1$, $P_2$, …, $P_J$. However, these probabilities will sum to 1: $P_1 + P_2 … + P_J = 1$.

For unordered qualitative variables (also known as polytomous variables) such as aircraft choice by the airlines, categories must be truly nominal and mutually exclusive.[11] Furthermore, the ordering of the numerical values of the variables is also of no importance.[12] Therefore, any category can be used as the baseline category. However, such choice is usually based on some a priori theoretical or operational motivation.

From equation (12), for j > 2, the probability distribution function can be generalized as follows:

(13) $P_n(i) = \dfrac{e^{\mu\beta' x_{in}}}{\sum\limits_{j} \in \mathcal{C}_n \ e^{\mu\beta' x_{jn}}}$  $\qquad \forall i \in C_n$

where i = i-th choice belonging to the complete set of choices, $C_n$. When j = 2, equation (13) reduces to equation (12), i.e., binary logit (0 and 1 being a special case). Furthermore, equation (13) defines a proper probability mass function since $\forall$ i $\in C_n$,

(14) $0 \leq P_n(i) \leq 1$

and,

(15)  $\Sigma_{j \in C_n} P_n(i) = 1$

That is, the probability of individual choices, $P_n(i)$ is positive (equation 14) and they sum to 1 (equation 15). Furthermore, all disturbances in the choices are assumed to be (i) independently-distributed, (ii) identically-distributed, and (iii) logistically-distributed. (i) – (iii) together are also known as iid property (Ben-Akiva and Lerman 1984).

The maximum likelihood function $(\ell^*)$ for a generalized multinomial choice model is given by the following equation:

(16)  $\ell^* = \prod_{n=1}^{N} \prod_{i \in C_n} P_n(i)^{y_{in}}$

where $y_{in} = \{1, 0\} = \{$if observations n chose alternative 1; zero otherwise$\}$. Equation (13) describes a logit for which parameters are linear corresponding to equation (16). Taking the logarithm of equation (16), we seek to attain maximum of $\ell^*$ as follows:

(17)  $\ell^* = \sum_{n=1}^{N} \sum_{i \in Cn} y_{in} \left( \beta' x_{in} - \ln \sum_{i \in Cn} e^{\beta' x_{jn}} \right)$

Taking first-order derivative of $\ell^*$ with respect to coefficients $(\beta_k)$ and setting it equal to zero, the first-order conditions are as follows:

(18)  $\sum_{n=1}^{N} \sum_{i \in Cn} \left[ y_{in} - P_n(i) \right] x_{ink} = 0 \ \forall \ k = 1,...., K$

or,

(19)  $1/N \sum_{n=1}^{N} \sum_{i \in Cn} y_{in} x_{ink} = 1/N \sum_{n=1}^{N} \sum_{i \in Cn} P_n(i) x_{ink}$

$$\forall \ k = 1,...., K$$

The estimator of $\beta_k$ that maximizes the above function $\ell^*$ is consistent, asymptotically normal, and asymptotically efficient (McFadden 1974). Equation (19) states that the average value of attributes for the chosen alternatives (left hand side of the equation) is equal to the average value predicted by the estimated choice probabilities (right hand side of the equation) (Ben-Akiva and Lerman (1984). If alternative-specific constant, for example, is defined for a particular alternative i, then, at the maximum-likelihood estimates, $\ell^*$, equation (19) reduces to the following:

(20)  $\sum_{n=1}^{N} y_{in} = \sum_{n=1}^{N} P_n(i)$

which implies that the sum of all choice probabilities for alternative i, taken over the sample, will equal the number in the sample that actually chose i. The estimated vector, $\beta_k$, is a vector consisting of slope parameters that will determine the effect of X vector on the probabilities of i-th choices. The computational methods and processes for solving the system of K equations in equation (20) are identical to those used in the binary logit case described earlier.

The empirical framework consists of six qualitative choices: Category 1 – Category 6 (Table 1). There were 92 and 87 distinct equipment types in 2003 and first two quarters of 2004, respectively. Reducing all these observed choices into six categories was necessary to minimize the computational cost and to derive reasonable results that are both manageable and meaningful.[13] Although relatively smaller aircraft are less important in terms of total enplanement, single engine piston aircraft and turboprops in particular, they are included to make the categories exhaustive. The detailed categorization together with some representative equipment types are provided in Table 1.

The X vector, i.e., vector containing exogenous variables, consists of the following variables: passengers in flight segments, stage length distance of segments, hub status of origin and destination airports, season of the year, and types of network. We consider the levels of passengers as an exogenous variable although passenger levels are also determined via a set of other exogenous variables, i.e., fares, income, population (Bhadra 2003). However, examining all these relationships falls outside the scope of this model. Furthermore, in the absence of cost data by equipment type for each flight segment, distance is used as a proxy in the empirical framework. It is hypothesized that distance affects cost positively; however, at a diminishing rate (Hoffer, Dresner, and Windle, forthcoming 2006).

**Table 1: Categorization of Scheduled Aircraft in the U.S. NAS**

| | Types of Aircraft | Representative Types of Equipment in this Category | Average Trip or leg Distance (miles) | Average Size Range | |
|---|---|---|---|---|---|
| | | | | Passengers (minimum) | Passengers (maximum) |
| Category 1 | Single Engine Pistons | CES-150 - 185; AERO-200; PA 12-32 | 150 | 3 | 20 |
| Category 2 | Turbo Props | ATR-42/72 Aerospatial; Beechcraft 1900; Dornier 328 Turbo; | < 250 | 30 | 37 |
| | | TurboProp 1-2 engine; JETST-31 BAE; JETST-41 BAE; | < 250 | 60 | 72 |
| Category 3 | Regional Jets (RJs) | Canadair RJ-100/R;Canadair RJ145-200; | 250-500 | 45 | 70 |
| | | Embraer EMB-135; Embraer EMB-145; EMB-140 | 250-500 | 45 | 70 |
| Cateogry 4 | Short-Haul Narrow Bodies | Boeing B-737-500; Boeing B-737-400; | 500-750 | 127 | 155 |
| | | Boeing B-737-300; Boeing B-737-100; | 500-750 | 105 | 129 |
| | | Boeing B-737-200C; Douglas DC-9-10; | 500-750 | 62 | 76 |
| | | Douglas DC-9-30; Douglas DC-9-40; | 500-750 | 87 | 107 |
| | | MD-80 & DC-9-80; MD-90-30/50; | 500-750 | 135 | 163 |
| | | Douglas DC-9-50; Boeing B-727-100; | 500-750 | 113 | 139 |
| Category 5 | Long-Haul Narrow Bodies | Boeing B-737-800; Boeing B-757-200; | 750-1500 | 155 | 189 |
| | | Euro Airbus A320; | 750-1500 | 131 | 161 |
| | | Boeing B-737/LR Boeing B-737-700/; Boeing B-727-200; | 750-1500 | 132 | 162 |
| Category 6 | Wide Bodies | Douglas DC-10-10; Douglas DC-10-30; | 1500-3000 | 278 | 340 |
| | | Douglas DC-10-40; Boeing B-747-100; | 1500-3000 | 256 | 312 |
| | | Boeing B-747-200; Boeing B-747-400; | 1500-3000 | 321 | 393 |
| | | Boeing B-767-200; Boeing B-767-300; | 1500-3000 | 158 | 194 |
| | | Boeing 777; Lockheed L-1011-1; | 1500-3000 | 239 | 293 |
| | | Lockheed L-1011-50; Douglas MD-11; | 1500-3000 | 299 | 379 |
| | | Euro Airbus A-300; Euro Airbus A310; | 1500-3000 | 205 | 251 |

Notes: For information relating to distance, and size in terms of passengers, see Aviation Week and space Technology (2003).

U.S. air transportation is heavily dependent on a hub-and-spoke type of network (Bhadra and Texter 2004). Available estimates indicate that more than 90% of scheduled passengers pass through some form of hub, i.e., large, medium, or small, while almost 70% of scheduled aircraft operations take place in a relatively large hub (Bhadra 2004). Consequently, it is likely that type of hubs may have some impact on type of aircraft that will be chosen to fly a particular segment. Furthermore, the choice of aircraft along with positioning or sequencing may also depend on the types of network an aircraft is serving. For example, while a regional jet (RJ) is likely to serve hub and spokes, a shorter haul narrow body may be used to serve medium to large hubs that are within medium distances. On the other hand, long-haul narrow body aircraft are likely to serve longer hauls and large hubs. Therefore, types of network, i.e., hub-to-hub and hub-to-spokes or between point-to-point, may influence the choice of aircraft as well. Finally, air travel has well-defined seasonality. While it is not clear why different aircraft will be chosen to serve different times of the year, a priori, there is no reason to reject that hypothesis either. Therefore, we empirically postulate that seasons of the year, high and low, may also impact the choice of aircraft.

The empirical framework can be specified as follows:

$$(21) \quad \begin{aligned} P_i\,(y_i \quad j\mid x_i, \beta) \\ (j = 1, 2, \ldots, 6) \end{aligned} \quad \begin{aligned} &= \alpha_{ij} + \beta_1\,(\text{passengers}) + \beta_2\,(\text{distance}) \\ &\quad + \beta_3\,(\text{OriginHubDummy}) + \beta_4\,(\text{DestinationHubDummy}) \\ &\quad + \beta_5\,(\text{network}) + \beta_6\,(\text{season}) + \varepsilon_i \end{aligned}$$

where passengers are the segment-pair passengers and distance is the stage length distance between two segment points. The probability of aircraft choice, independent of the size or types, would likely be positively influenced by the number of passengers and distance because the more passengers there are the more likely that aircraft will be selected for some of the trips. The greater the distance, the more likely passengers will select air rather than alternative modes of travel. However, the higher the stage length, the more likely the choice will be a relatively larger aircraft (e.g., narrow body) rather than smaller types (e.g., turbo props) within the general choice of aircraft.

The OriginHubDummy and DestinationHubDummy are dummy variables representing the airports if they were large airports[14] from where the segment flight originated or landed. The status of origin and destination airports plays important roles in choice of aircraft. While an airport size may influence the aircraft choice, the size or type of aircraft depends on the types of airports at both the origin and destination. For example, while narrow body aircraft may likely be chosen for flights between two large hubs, regional jets are often flown in hub-and-spoke networks, where a large hub distributes passengers to smaller spoke airports.

To capture the characteristics of networks, four variants of the airline network variable have been used in the analysis and are defined in the following fashion: point-to-point (PP) is defined as air travel that takes place between small and non-hub airports; hub-to-hub (HH) is defined as air travel that takes place between two major hubs. Hub-to-spoke (HS) is outbound or hub-to-spoke travel, and SH is defined as air travel from spoke-to-hub or inbound traffic. With definition of these four variants of the network, the network dummy (k = 0, 1, 2, 3) is defined as point-to-point (PP; k=0), hub-to-hub (HH; k=1) and hub-to-spoke or outbound traffic (HS; k=2) and spoke-to-hub or inbound traffic (SH; k=3). The sign of this dummy variable is not definite a priori because it may be influenced by factors (e.g., flight frequencies) that are outside the proposed model. Nonetheless, both the sign and the magnitude estimated from this dataset may shed light on the types of aircraft choice.

Finally, it is hypothesized that choice of aircraft may be affected by seasons. Empirically speaking, air travel goes through seasonal variations, peaking during spring and summer (i.e., April-September; season dummy = 1) and hitting a trough during fall and winter (i.e., October-March; season dummy = 0). Given the data for only the first two quarters of 2004, January-March has been used with season dummy = 0 and 1 otherwise (April-June). Hence, we expect this sign to be positive a priori.

The maximum likelihood (ML) estimation procedure is used for estimating equation (21). There are two reasons for which ML is often chosen as a general approach for estimating logistic regressions, especially for large samples. First, ML estimators are consistent, asymptotically efficient, and asymptotically normal. Second, it is fairly straightforward to derive ML estimators. These are desirable properties given that large samples are used in the empirical analysis [(SAS/ETS version 8 (1993); and Allison (2001)].

## DATA

Data for this exercise comes from the Bureau of Transportation Statistics/Department of Transportation's (BTS/DOT) T100 schedule. T100 is the transportation schedule of the Form 41 data that every scheduled carrier is now required to submit to the DOT every quarter. T100 is broken into two parts: T100 market segment which covers all the O&D markets, as opposed to segments;[15] and the T100 segment which provides data for market segments. T100 segment is the Data Bank 28DS of Form 41 which provides traffic, capacity, and aircraft equipment used by airlines in the segments they served (DOT 2001). The data are reported by scheduled air carriers operating non-stop between airports located within the boundaries of the United States and its territories (for availability of data, see U.S. DOT 2004).

T100 segment data can be best explained using the diagram in Figure 4.

For the empirical analysis reported in this paper,[16] data for two time periods: Quarter 1 and 2 of 2004 is used. Of the total 140,647 observations or segments that were distinctly

**Figure 4: Segment and Market Travel**

**Flights from Los Angeles (LAX) to Salt Lake City (SLC) to Denver (DEN)**

| LAX | SLC | DEN |
|---|---|---|
| 100 enplane | 40 deplane; 50 enplane. | 110 deplane |

**Non-Stop Segments:** Represented by straight arrows above, i.e., number of passengers transported between points.

        LAX to SLC: 100 passengers transported;

        SLC to DEN: 110 passengers transported;

**On-Flight Markets:** Represented by curved lines above, i.e., where passengers are enplaned and deplaned on a <u>flight</u> (flight number).

        LAX to SLC: 40 passengers;

        LAX to DEN: 60 passengers;       For a one-stop flight, the number of passengers would

        SLC to DEN: 50 passengers;        be the same number under segment and market.

Source: DOT (1992)

reported in these two quarters, there were 2,374,148 departures in the first quarter serving 147.96 million passengers. In the second quarter, there were 2,500,030 departures serving 168.75 million passengers. Thus, total departures and total passengers for the entire sample were, respectively, 4,874,178 and 316.71 million. The number of observations, total departures, and total passengers by aircraft categories are reported in Table 2.

**EMPIRICAL RESULTS**

The multinomial logit procedure was used to estimate equation (21). Multinomial logit models use maximum-likelihood estimation for polytomous dependent variables,[17] and hence it is also known as polytomous logistic regression. Notice here that the groups formed by the categories of a polytomous dependent variables are not truly independent (i.e., choice of one aircraft in a segment may also depend on other aircraft choices as well) thus preventing one from simply doing as many separate logistic regressions as there are categories. Multinomial logit handles non-independence by estimating the models for all outcomes simultaneously except, as in the use of dummy variables in linear

regression, one category is used as a baseline. Since effects must sum to zero, the model for the reference group can be reproduced given the other parameters. For the estimation, single engine piston aircraft (Category 1) is used as the baseline. This category is chosen as the baseline because it serves as the lowest category in cardinal ranking, evaluated both in terms of aircraft size and average haul. Therefore, all other categories can be thought of as a cardinal upgrading over Category 1 aircraft choice.

Results from the estimation have been summarized in Table 2. It is important to note that interpretation of the coefficient values is not the same under qualitative choice models as they are under linear and many non-linear models. It is complicated by the fact that estimated coefficients, i.e., effect coefficients, from an MNLM model cannot be interpreted as the marginal effect on the dependent variable. Nonetheless, their signs and magnitudes provide important information. Estimated effect coefficients, for example, represent the change in the log odds of the dependent variable, i.e., a particular type of aircraft choice due to changes in the explanatory variables. Despite the difficulties in explaining estimated coefficients directly, positive values of $\beta_i$ would imply that

**Table 2: Logistic Regression Results for Aircraft Choice (first two quarters of 2004)**

| Aircraft Category (1) | No. of Obs (counts or segments) (2) | Total Departures (3) | Total Passengers (4) | Intercept (5) | Passengers (6) | Distance (7) | Large Origin Airport Dummy (8) | Large Destination Airport Dummy (9) | Network Dummy (10) | Season Dummy (11) |
|---|---|---|---|---|---|---|---|---|---|---|
| Single Engine Pistons: Category 1 | 8,829 (6.28) | 64,191 (1.32) | 110,878 (0.04) | -- | -- | -- | -- | -- | -- | -- |
| Turbo Props: Category 2 | 30,811 (21.91) | 939,562 (19.28) | 15,098,642 (4.77) | -0.3589* (188.31) | 0.0146* (674.48) | 0.0098* (1372.43) | 5.1750 (0.0873) | 5.5735 (0.0450) | -0.1485 (0.0003) | 0.2677* (98.73) |
| Regional Jets: Category 3 | 29,540 (21.00) | 1,358,487 (27.87) | 45,937,594 (14.50) | -4.1243* (10840.81) | 0.0147* (685.38) | 0.0167* (3759.84) | 6.8390 (0.1524) | 6.9908 (0.0708) | 0.2347 (0.0007) | 0.1616* (23.24) |
| Short-Haul NBs: Category 4 | 34,443 (24.49) | 1,392,810 (28.58) | 121,634,160 (38.40) | -3.8153* (10821.04) | 0.0149* (702.90) | 0.0182* (4438.09) | 6.2784 (0.1285) | 7.0880 (0.0728) | -0.3958 (0.0020) | 0.2814* (70.06) |
| Long-Haul NBs: Category 5 | 33,299 (23.68) | 1,012,367 (20.77) | 114,068,283 (36.01) | -5.2653* (16015.09) | 0.0149* (700.87) | 0.0194* (5031.32) | 6.6669 (0.1449) | 7.5460 (0.0825) | -0.4080 (0.0022) | 0.2398* (47.12) |
| Wide Bodies: Category 6 | 3,725 (2.65) | 106,761 (2.19) | 19,882,397 (6.28) | -10.3819* (4293.65) | 0.0149* (704.81) | 0.0202* (5447.33) | 7.5829 (0.1874) | 8.3659 (0.1015) | -0.2241 (0.0007) | 0.2503* (26.66) |
| N = | 140,647 (100) | 4,874,178 (100) | 316,731,954 (100) | | | | | | | |

| | |
|---|---|
| Max-Scaled R² | 0.6865 |
| Likelihood Ratio | 151,467 |
| Score | 111,104 |
| Wald Statistic | 50,409 |

Notes: Category = 1 (i.e., Single Engine Piston Aircraft) has been used as reference category. '*' represent statistical significance at 99% levels of confidence. Wald-Chi squares are given underneath the estimated parameters. NBs: Narrow Bodies.

increasing $\beta_i$ will increase the probability of selecting a particular aircraft type. As noted earlier, the estimated parameters $\beta_k$) hold for individual choices, estimated over the entire sample maximizing log-likelihood function ($\ell$). Like their counterparts in linear models, estimated $\beta_k$ from multinomial logit models represent, on average, the sample as well.

Estimated parameters in the model (Table 2) indicate that passengers, stage distance, large hub airports and high season - all have positive impacts on the odd ratios of all aircraft choices. This is expected because air travel is facilitated by these factors, i.e., aggregation of passengers in a relatively large airport wanting to fly a distance that is often outside the range of driving.

The network dummy, however, affects (statistically insignificant) the choices negatively for all cases other than Aircraft Category 3 (RJ). Thus, as the network becomes more hub-oriented (dummy assuming value from 0 to 3), the likelier that RJs will be chosen. This is perhaps representative of the fact that RJs play crucial roles in performing hub-and-spoke activities throughout the NAS.

More importantly, estimated coefficients for passengers are monotonic in aircraft choices because aircraft choices are ordered according to their sizes.[18] Hence, it is expected that the estimated parameters for passengers would be positively monotonic in aircraft choices. That is, the larger the aircraft the larger the estimated coefficient. A similar a priori empirical hypothesis with respect to distance is also confirmed from the estimated coefficients. The larger the aircraft, the larger the estimated coefficients for distance. Although hubs and network are positively related to the aircraft choices, none of the airports hub coefficients is statistically significant. The same is true for the network dummy. The estimated seasonal variable coefficients indicate that all aircraft types are more likely to be selected in the second quarter with varying degrees of impact on aircraft choices. It is interesting to note that the busier season tends to have a relatively smaller positive impact on the choice of regional jets. Results for other aircraft choices are not so obvious with respect to the seasonal dummy.

Finally, we have estimated the model for predicting the lowest value of the dependent variable. In other words, the estimated model predicts the probability that the aircraft category choice is equal to 0. However, the SAS procedure allows the reverse, i.e., predicting the highest value (equal to 1), by specifying the 'descending' option in the model statement.

Despite the lack of statistical relevance of some of the variables, the above specification was retained in making aircraft choices for two reasons. First, the ultimate reason for using the above choice model is to replace arbitrary assumptions with respect to load factor and aircraft size for determining aircraft operations by segment pairs. These aircraft operations are driven, among other things, by passenger flows among commercial airports (Bhadra et. al. 2005). Hence, more information, i.e., types of airports and networks in particular, are beneficial in deriving these aircraft operations. Second, and perhaps more importantly, the results reported here may be somewhat biased by the choice of sample. Following the restructuring that occurred after 9/11, the airline network appears to have become more point-to-point oriented over time (Bhadra and Texter 2004). Hence, larger hubs and variants of hub-to-spoke networks have become relatively less important than they were prior to restructuring. It is not evident, however, that these changes are permanent. Given the uncertainty which is typical of a transitory time, it was decided to keep both the large hubs and network dummy in the above specification.

Notice also that $\beta$s are the estimated parameters for each alternative aircraft choice. These point estimates are critical in determining the probability of the choices. However, they are estimated for the entire individual samples of alternative choices, i = I, 2, ….6. The estimation procedure allows joint maximization of likelihood estimator ($\ell$). Therefore, the $\beta$s perform well for the entire sample and ought not to be evaluated for each observation point.

Overall model results indicate that the model specification is indeed robust. Results testing the null hypothesis that all explanatory variables have coefficients equal to zero (0) prove to have been conclusively rejected.[19]

Wald-Chi square estimates (reported under the estimated parameters), accompanied with probability values less than .01, indicate that at least one of the coefficients, if not all, is not 0. In other words, the model has a good overall fit. The Wald chi-squares are calculated by dividing each coefficient by its standard error and squaring the result.[20]

While estimated parameters provided an assessment of the overall validity of the specified model for the entire sample, there are two additional criteria by which we can judge the performance of the model. First is the estimated log odd ratio[21] corresponding to the explanatory variables.[22] For example, the passengers variable for the Aircraft Category choice = 2 (i.e., turbo prop) has a point estimate of odd ratios of 1.015 with 95% Wald confidence limits of 1.014 – 1.016. This implies that the predicted odds of choosing a turbo prop increases by about 1.5% with a one-unit increase in passengers. While passenger is an important explanatory variable, the monotonic relationship is most evident with the distance variable; with a unit increase in distance (i.e., a mile), predicted odds that a higher category aircraft will be chosen increases monotonically. In particular, for the Aircraft Category choice = 2 (turbo props), the point estimate of odd ratios with respect to distance is 1.010 (with confidence interval of 1.009 – 1.010). For Aircraft Category choice = 3 (regional jets), the point estimate of odd ratios is 1.017 (with confidence interval of 1.016 – 1.017); for Aircraft Category choice of 4 (short haul narrow body), it is 1.018 (with confidence interval of 1.018 – 1.019), 1.020 for Category 5 (long haul narrow body) with a confidence interval of 1.019 – 1.020, and 1.020 for wide body with confidence interval of 1.020 – 1.021.

Although the model is estimated on the sample as a whole, and hence, estimated parameters are valid for the sample, often, results of the Logit Choice model are evaluated at each observation point. Evaluating the above model at each observation point and comparing the observed occurrences with that of estimated probabilities may reveal further information regarding the structure of the model and hence the underlying choices. The estimated parameters can be used to predict the aircraft choice responses at the individual observations to evaluate the model's performance. These predicted responses were compared to that of actual choices. Actual choices have been reported in column 2 of Table 3. When predicted responses matched exactly to that of the actual, they are called "exact" predicted responses as reported in Table 3, column 3. The numbers in column 3 are the number of observations in which the model correctly predicted the actual aircraft chosen. The numbers in parentheses represent the percent of observations that duplicated the actual choice.

Overall, the choice model seems to perform well for aircraft choice categories 2, 3, and 5. For the single engine piston aircraft or, AC category 1, the model predicted very few choices correctly (i.e., 1,323, or, 14.98% of the total 8,829 observations) while for the wide bodies or Aircraft Category 6, the model predicted only 3.49% (or 130 out of 3,725 actual choices). For turbo props, category 2, the model's prediction matched actual choice in about 74% of the cases (22,699 cases out of 30,811 observations). For regional jets (category 3), the model's prediction matched actual choice in about 60% of the observations (17,702 out of a total pool of 29,540). For short-haul narrow bodies, the model could replicate actual choice only 32% of the time (11,123 out of a total of 34,443). For long-haul narrow bodies (Category 5), the model could predict actual choice about 57% of the time (19,084 of the total of 33,299). The overall performance of the model was 51% (72,061 out of a total sample of 140,647 actual choices).

There are a couple reasons for the relatively poor choice prediction performance for the smallest and largest size aircraft. Many of the wide bodies in the system serve the international routes most frequently. Many of the domestic segments they fly are for positioning purposes for international routes rather than to serve domestic segments exclusively. Single-engine piston aircraft are not commonly used for most scheduled air transportation. Wherever they are in use, there are other factors, e.g., codesharing with particular regional partners with single-engine piston aircraft in the inventory serving

**Table 3: Performance of the Model: Actual vs. Predicted**

| AC Category (1) | Actual Choice (2) | "Exact" Predicted Response (3) | "One-Off" Predicted Response (4) | Exact plus One-off Predicted Responses (5) |
|---|---|---|---|---|
| | 2004: Q1 – Q2 | Predicted response = actual choice | Predicted response is off one category of the Actual | (3+4) |
| **Single Engine Pistons: Category 1** | 8,829 (6.28) | 1,323 (14.98) | 7,500 (84.95) | 8,823 (99.93) |
| **Turbo Props: Category 2** | 30,811 (21.91) | 22,699 (73.67) | 7,260 (23.56) | 29,959 (97.23) |
| **Regional Jets: Category 3** | 29.540 (21.00) | 17,702 (59.93) | 8,817 (29.85) | 26,519 (89.77) |
| **Short-Haul NBs: Category 4** | 34,443 (24.49) | 11,123 (32.29) | 19,939 (57.89) | 31,062 (90.18) |
| **Long-Haul NBs: Category 5** | 33,299 (23.68) | 19,084 (57.31) | 6,080 (18.26) | 25,164 (75.57) |
| **Wide Bodies: Category 6** | 3,725 (2.65) | 130 (3.49) | 2,788 (74.85) | 2,918 (78.34) |
| **N / average % of correct response** | 140,647 (100) | 72,061 (51.24) | 52,384 (37.25) | 124,445 (88.48) |

Notes: It is evident that choice = 2 - 5, in effect, have 2 grouping possibilities, one higher and/or one lower. Thus, one-off predicted response for short-haul narrow body (Aircraft Category = 4), for example, may have two grouping possibilities, which are Category 3 (regional jets) or Category 5 (long-haul narrow body). In comparison, single engine piston aircraft and wide bodies have only one grouping possibility, turbo props for the former and long-haul narrow bodies for the latter.

small markets that may be responsible for these choices. In other words, there are factors outside this generalized model that can explain choice behaviors for the largest and smallest aircraft.

When the exactness is made somewhat flexible, and one choice ± from the actual choice was allowed, predicted responses that are "one-off" are obtained (reported in column 4 of Table 3). For example, when the choice of an aircraft (say, for example, Aircraft Category = 4) can have a value of both 3 and 5, in addition to its

exact value of 3, this is called the "one-off" predicted response. The "one-off" category was created in order to account for the fact that often the choice of aircraft is not nearly as distinct as implied by the categorical cardinal choice of 1, 2, …, 6. That is, this one-off possibility allows one to explicitly account for the fact that many of the choices are somewhat similar (most obvious being short haul and long-haul narrow bodies, for example), and therefore, may have some continuity as opposed to discrete choices

resulting from the MNLM. Furthermore, this flexibility allows the model to be more useful for operational use. Column 4 of Table 3 summarizes these results.

As expected, flexibility results in better predictive performance in some cases. The gain from this one-off allowance is obviously largest for which the exact match is poor. Thus, the predictive performance gain in the smallest and largest aircraft choices is highest with somewhat higher gain recorded for short-haul narrow bodies as well.

The last column of Table 3 aggregates results of exact and one-off matches. As evident, the estimated model with one-off allowance is capable of predicting almost nine out of 10 actual aircraft choices from the sample.

Notice that the above comparison between "exact" and "exact and one-off choices" reveals some important information regarding the underlying structure of the choice modeling. Recall that MNLM assumes that errors are identically and independently distributed. That is, the variance is constant across the choices and these choices are also independent choices. A comparison of the predictive choices under exact (column 3) to that of exact plus one-off choices (column 5) reveals that this assumption may not be true. For example, by allowing for one-off choices, the maximum gain occurred to wide bodies (i.e., over 22-fold increase or (78.34%/3.49%)), single-engine piston (almost 7-times; from 14.98% to 99.93%) and short-haul narrow bodies (almost three times). Thus, combining choices reveal that maximum gain is attained for choices for which the "exact" prediction is rather poor. In other words, there is a possible prediction gain to be had by grouping these choices with their neighbors, i.e., single-engine pistons with turbo-props while combining wide bodies with long-haul narrow bodies and short-haul narrow bodies with regional jets.

**CONCLUSION**

In this paper, a multinomial logistic regression model was used to determine the choice of aircraft in the U.S. NAS. By categorizing all aircraft into six categories, it was found that passengers, distance, types of airport hubs, network, and seasons are capable of estimating these choices fairly well. The findings indicate that the estimated model is capable of predicting five out of 10 of these choices exactly (51%) for the total sample, with relatively better performance for turbo props, regional jets, and long-haul narrow bodies. The model performs poorly for the smallest and largest of the aircraft. Almost nine out of 10 aircraft choices can be explained by the model if "one-off" predicted responses are allowed in place of exact predicted response.

These findings have important implications. First, the estimated model enables mapping of passengers onto aircraft choices, given distance, status of hubs, types of network, and seasons. This provides another tool that can be used to replace arbitrary assumptions of load factors and aircraft size. Second, and most importantly, the empirical relationship between passengers and aircraft choices allow derivations of aircraft operations by market segments. This further allows airlines to generate schedules or timetables specific to airports that are driven by, among other things, passenger forecasts. This ensures that we can model and simulate the operations of U.S. NAS far more efficiently, corresponding to different passenger-demand scenarios than was previously the case.[23]

There are quite a few areas of future research. Allowing flexibility in choices reveals that there is a possible gain to be had by grouping (or *nesting*) the choices with their adjacent neighbors, e.g., single-engine pistons with turbo props while combining wide bodies with long-haul narrow bodies and short-haul narrow bodies with regional jets. Modeling these nests may lead to improvement in the results. Additionally, the data can be segmented by distance categories, i.e., short haul ($\leq$ 750 miles), medium haul (750-1500 miles), and long haul ($\geq$ 1500 miles) to capture the effects of types of markets on aircraft choices. This may improve the results because aircraft choices will incorporate weighted distances. This added information may benefit the estimation substantially. One of the many

shortcomings of the above model is that it does not consider airline behavior in aircraft choice. By incorporating airline-specific behavior explicitly (i.e., either by specifying behavioral equations, or at the least, via dummy variables), the model can be improved. Finally, passenger demand that is determined by economic factors can be modeled and used as determinants of aircraft choice as well. These are the tasks for future research.

**Endnotes**

1. A passenger getting on board is counted as an enplanement. This may result in completion of a trip, e.g., origin at one end and the destination on the other (O&D), leading to enplanement and O&D passengers yielding the same count. If, on the other hand, a trip has a stop-over or a connection somewhere else other than O&D, enplanement will be higher than the O&D passengers count.

2. Delta's then-Chairman Leo Mullin brought this point home clearly when he commented on 9/11/2003 "…matching of market to the right size of aircraft is a crucial ingredient to our success" retrieved from http://www.enquirer.com/editions/2003/09/11/biz_911leo11 html.

3. An important distinction between actual and scheduled should be made at this point. In airline operations planning, scheduled or planned aircraft positioning optimizes inventory planning, resulting in cost minimization and/or perhaps revenue maximization. In this exercise, however, we restrict the analysis to actual aircraft choice, rather than planned choice.

4. At the end of 2002, Southwest had 27 B-737/200, 194 B737/300, 25 B737/500 and 129 B-737/700. Even within this simplified fleet structure, four choices (i.e., B737/200 – 700) within the broad category of short-haul narrow body makes the choice of aircraft within a market or in a segment complicated.

5. The six major airlines are American, United, Delta, Continental, US Airways, and Northwest. Their joint market share is around 70%, at present. However, only very recently (May-June, 2003), Southwest has become the largest domestic airline carrying more passengers than any of the top six carriers. For example, Southwest carried 6.5 million domestic passengers during the month of May, 2003, beating Delta Air Lines which carried 6.3 million passengers, and American Airlines carrying 6.2 million passengers (see for details, http://www.dfw.com/mld/dfw/business/6518020. htm; retrieved, August, 12, 2003).

6. While a relatively high proportion of Southwest's passengers use point-to-point service, almost a quarter of them make connections as well.

7. Indirect profit is derived by substituting the optimal input demand functions from the first order conditions. Because input demand functions are expressed in terms of input and output prices and fixed input, hence, the indirect profit function is dependent on them as well. The indirect profit function is used to evaluate the optimality of the profit.

8. This section is developed for demonstration purposes. The binary logit model is rather simplistic because the majority of the airlines have more than one type of aircraft. Nonetheless, binary choice logit provides a conceptual framework that is relatively easy to understand.

9. Instead of strictly defining one category of aircraft, one can also put all others in one category. For example, choice of one category (narrow bodies) and all others (i.e., all non-narrow bodies) can be defined as binary choice.

10. Binary choices have been widely discussed in the literature, primarily to explain voting behavior (see Pindyck and Rubinfeld (1991) for a theoretical framework; and, http://www2.chass.ncsu.edu/garson/pa765/logit.htm for applications in voting behavior context).

11. For example, a category called "lowest aircraft" cannot be used because it is not truly nominal. Instead, a category representing "lowest," however defined, should be used. Similarly, aircraft categories i (short-haul narrow body), j (long-haul narrow body), and k (overall narrow body) together can not be used because i, j, and k are not mutually exclusive.

12. In other words, assignment of numerical values to a particular category and ordering do not have any importance. Aircraft category i = cessnas and pipers and aircraft category j = turbo props is the same as the opposite assignment (i.e., aircraft category j = cessnas while aircraft category i = turbo props).

13. As we will see in our discussion of the performance of the model, statistical results depend on our classification of categories. Based on the criteria described in Table 1, the categorization reduces numerous aircraft categories into six broad categories. Despite careful categorization, errors can occur in both defining which aircraft belongs to what category as well as accounting. Therefore, caution should be used in interpreting the results of this paper.

14. Airport hubs are defined in two ways. First is in terms of total enplanements (i.e., physical magnitude), as defined by Department of Transportation (DOT), Federal Aviation Administration (FAA). Under this definition, there are four kinds of airports: large ($\geq$ 1% of total U.S. enplanements), medium (0.25%-0.999% of total U.S. enplanements), small hubs (0.05%-0.249% of total U.S. enplanements); and nonhub (< 0.05 of total U.S. enplanements). The second definition categorizes an airport as a hub where a major commercial air carrier has more than one bank structure as a hub, i.e., operational or functional definition. Under this definition, an airport is defined as a hub where inbound flights are scheduled to arrive from multiple origins within a short space of time thus creating a bank of passengers. The coordinated arrival and departure banks together form a wave. There is some empirical correspondence between physical and operational hubs. Hypothetically speaking, an airport can be an operational hub without being a physical hub (e.g., airports serving only connecting passengers); while a physical hub may exist without being an operational hub (i.e., all origin and destination passengers).

15. Segments are the connecting flights in an O&D trip. For example, an O&D trip between Los Angeles (from LAX airport) to Denver (to DEN airport) may have a connection at Salt Lake City (at the SLC airport), giving rise to two segments for the same O&D trip. That is, for a LAX-DEN O&D trip, LAX-SLC and then SLC-DEN are the two segments of the same trip.

16. Larger empirical analysis consists of 18 quarters of data: Q1: 2000 – Q2: 2004 with more than 1.3 million records. To limit the analysis to a manageable magnitude and also to put more emphasis on current quarters, results are reported from two representative quarters, 1st and 2nd quarters of 2004. However, the aggregate relationships reported here do not change much when estimated with larger datasets.

17. Polytomous variables are also known as unordered qualitative variables, such as aircraft choice by the airlines. The ordering of the numerical values of the variables as such has no importance. Notice also that these categories must be truly nominal and mutually exclusive.

18. It is important to recognize here that ordinal ranking of aircraft (i.e., aircraft choices of 1, 2, …,6), which is somewhat arbitrary, do not exactly correspond to cardinal sizes of the aircraft. In

other words, aircraft choice of short-haul narrow body (Category 4), for example, is not exactly double the size of turbo props (Category 2). The same reasoning applies, with lesser extent, to distance as well.

19. This hypothesis is tested by an overall F-test in a linear regression.

20. Without squares, these estimates are the same as t or z statistics. The p-values calculated from a normal table would be exactly the same as the Chi-Square p-values reported.

21. The odds are a familiar way of representing probability. It is defined as the ratio of the probability that the event of interest occurs to the probability that it does not and is often estimated by the ratio of the number of times that the event of interest occurs to the number of times that it does not. Log odds ratio estimates the effect of explanatory variables (Xs) on the ratio of the probability that one choice (over others) will be predicted. The odds ratio is obtained by raising the value of the parameters, βs to their exponents (i.e., exponentiating) associated with the explanatory variables (Xs).

22. Available from the author, upon request.

23. For an application of this procedure in generating schedules, see Bhadra, et. al. (2005).

**References**

Airline Monitor. "Block Hour Operating Costs by Airplane Type for the Year 2002," August, Part II, monthly issues, Ponte Vedra Beach, Florida, 2003.

Air Transport Association.  See http://www.airlines.org/public/home/default1.asp, 2004.

Allison, Paul D. *Logistic Regression Using the SAS System: Theory and Application*. SAS Institute, 2001.

Aviation Week & Space Technology. *Aerospace Source Book*, January 13, 2003.

Ben-Akiva, M. and S. R. Lerman. *Discrete Choice Analysis: Theory and Applications to Travel Demand*. The MIT Press, Cambridge, MA, 1984.

Bhadra, D., J. Gentry, B. Hogan, and M. Wells. "Future Air Traffic Timetable Estimator." *Journal of Aircraft* 42(2), (2005): 320-328.

Bhadra, D. and P. A. Texter. "Airline Networks: An Econometric Framework to Analyze Domestic U.S. Air Travel." *Journal of Transportation and Statistics* 7(1), (2004): 87-102.

Bhadra, D. "Demand for Air Travel in the United States: Bottom-Up Econometric Estimation and Implications for Forecasts by O&D Pairs." *Journal of Air Transportation* 8(2), (2003): 19-56 .

Fan, Terrence. "Smaller Aircraft for More Profits? A Preliminary Examination of Airlines' Fleet Size Decision With Fare and Demand Distributions." *Journal of the Transportation Research Forum,* published in *Transportation Quarterly* 56(3), (2002): 77-93.

Hofer, Christian, M. Dresner, and R. Windle. "Financial Distress and US Airlines Fares." *Journal of Transportation Economics and Policy* (forthcoming), 2006.

Hosner, D.W. and S. Lemeshow. *Applied Logistic Regression*. New York: John Wiley & Sons, 1989.

McFadden, D. "Conditional Logit Analysis of Qualitative Choice Behavior." P. Zarembka ed. *Frontiers in Econometrics.* New York: Academic Press (1974): 105-142.

Miller, B. and J-P Clarke. "Investments Under Uncertainty in Air Transportation: A Real Options Perspective." *Journal of the Transportation Research Forum* 44(1), (2005): 61-74.

Pindyck, R. S. and D. L. Rubinfeld. *Econometric Models and Economic Forecasts*, Third Edition. McGraw-Hill, New York, 1991.

SAS/ETS Software. "Applications Guide 2: Econometric Modeling, Simulation, and Forecasting." Version 6, First Edition, SAS Institute, Cary, NC, 1993.

SAS Institute Inc. *Logistic Regression Examples Using the SAS System*. Cary, NC, 1995.

SpeedNews. "Source for Aviation News and Information," see http://www.speednews.com/ 2003.

U.S. Department of Transportation. Bureau of Transportation Statistics (BTS), "Aviation Data", (2004) see http://www.transtats.bts.gov

U.S. Department of Transportation. *Code of Federal Regulations,* Ch II (1-1-01 Edition), Pt. 241 (2001), Washington, D.C., Office of the Secretary.

U.S. Department of Transportation. O&D Survey Reporting Regulations for Large Air Carriers: Code of Federal Regulations Part 241 (1992), Section 19-7, Office of the Secretary, Washington, D.C.

Yu, Junhua. "A Nested Logit Approach to Airline Operations Decision Process." Working paper, East Carolina University, 2003.

**Acknowledgement**

*Dipasis Bhadra is a Principal Economist with the MITRE Corporation's Center for Advanced Aviation system Development (CAASD), a federally-funded research and development corporation. At CAASD, Dipasis works in areas of quantitative modeling of air transportation system including passenger flows between airports, positioning of aircraft, and configuration of networks. Prior to joining CAASD in 2001, Dipasis worked at the World Bank evaluating feasibility of infrastructure projects in South Asia. With a Ph.D. in economics (1991) from the University of Connecticut, Dipasis also teaches introductory economics at the local community college, and holds professional positions at the Transportation Research Board (TRB) and American Institute of Aeronautics and Astronautics (AIAA).*